7 | Discovery Tools

Michael Courtney, Indiana University

WHAT IS DISCOVERY?

When S. R. Ranganathan provided the directive "save the time of the reader" (his fourth law of library science), he undoubtedly may not have predicted the significant shift in information distribution that is being fully realized in the 21st century as library collections have moved quickly to include (and in many cases replace print counterparts) a vast array of digital and electronic content (1931, p. 337). What Ranganathan fully understood, though, and why this simple phrase is still entirely relevant in spite of itself, is that it has been unwittingly responsible for library administrative reform ever since. Whereas his era was more concerned with shelf arrangement, catalog development, and the noticeable inclusion of standard bibliographies that would serve to uncover the library's resources to its users, contemporary libraries now face a new and complex dilemma—one that Lorcan Dempsey acknowledged as a shift in user focus (and need), wherein "the context of information use and creation has changed as it transitions from a world of physical distribution to one of digital distribution" (2012). Such a shift has necessitated the development of tools to aid in the discovery of resources and, to be sure, resource discovery is arguably one of the most pressing issues that concerns libraries today. Technological innovation in an increasingly online world has changed not only the availability of information resources, but also how those resources are being used and discovered. As Dempsey and others have accurately noted, changes in information format have altered access paradigms to the point that it is not often entirely

necessary to visit the physical library to consult resources. Rather, the focus has shifted from the localized information repository (the physical library) to the vast network of information that surrounds us in a virtual sense.

While the library has traditionally relied on the catalog, research guides, knowledgeable staff, and other finding aids, the arrangements of which are designed to orient the user to the information's organizational structure, the modern challenge that is presented concerns leveraging networked technology with new, more appropriate tools to facilitate resource discovery. The concept of discovery is not entirely new, however much the pragmatic approach of realizing information discovery has radically changed. Historically, librarians (and libraries) have developed tools all along to aid the user in identifying relevant information amongst an enormous pool of resources. Some parallels to our modern dilemma of information access can be made as far back as the 17th century, when more economical book production methods arose, creating an increase in information (and, in turn, knowledge) accessibility, as is the case today. The development of tools to aid in accessibility, from printed catalogs to bibliographies and beyond, gradually demanded classification systems as a means of collecting and cataloging growing collections, many of which are still employed in the present day. Former concerns of serendipitous information retrieval and a collection's browsability still resonate loudly centuries later. The concept of discovery of information across a network, however, has its roots in the classical field of information retrieval, the focus of which traditionally concerned relevant document searching across static, fixed, and typically text-based collections. In this model, the user is presumed to have a finite understanding of the information need insomuch as the user is able to accurately express the need to conduct a search within the system. In this approach, more contemporary concerns of the search process (e.g., information literacy and discovery) are theoretically ignored, limiting the effectiveness overall of the information retrieval system. Conversely, *information* or *resource* discovery, according to Clifford A. Lynch, "can be used to describe a complex collection of activities that can range from simply locating a well-specified digital object on the network all the way through lengthy iterative research activities . . . often involv[ing] the searching of various types of directories, catalogs, or other descriptive databases" and typically operating on "surrogates (such as descriptions) of actual networked information

resources" (1995, p. 1506). The terms information discovery and resource discovery, while having subtle, nuanced differences, are often used interchangeably in this regard.

Implicit in the concept of discovery is an effort by the user to locate (or, in effect, explore) the unknown. A user may have in mind characteristics or general knowledge of an overall information need, but here the emphasis is on locating resources that are unknown yet underscore the base need. From a library perspective, discovery tools allow users to not only discover "hidden" collections and uncover, serendipitously, new information relevant to the overall need, but also to facilitate navigation through library collections that have become increasingly complex and diverse in format. Such tools, which contemporarily have morphed from federated search capabilities (single query searching across multiple resources, or databases, the results of which are aggregated and displayed to the user, and typically referred to as metasearching) to discovery layers, web-scale discovery services, and beyond, will be the focus of the remainder of this chapter.

WEB-SCALE DISCOVERY

Web-scale discovery tools, or services, allow a user to "search seamlessly across a wide range of local and remote content and provide relevance-ranked results" and additionally to have "the ambitious goal of providing a single point of entry into a library's collections" (Breeding, 2014, p. 25). Perhaps more succinctly, Jason Vaughn suggests web-scale discovery "can be considered as deep discovery within a vast ocean of content" (2011, p. 5). The structure of a web-scale discovery tool employs a central index (or, variously, base index or unified index) of content that has been preharvested, from which a user can search across localized collections, open access resources, and subscription-based resources, all using a feature-rich discovery layer. Currently, the four major vendors of web-scale discovery tools that offer both a central index and a discovery layer are EBSCO (EBSCO Discovery Service), ProQuest (Serials Solutions' Summon), OCLC (WorldCat Local), and Ex Libris (Primo Central Index). The central index, as is in the case of the major vendors mentioned previously, is a collection of citations and full text from publishers, subscription databases, and open-source collections, as well as MARC records from library catalogs. Metadata from local collections that have been digitized, in addition to content from

institutional repositories, may further enhance the central index. The discovery layer, or the user interface, allows for interactive search and display of content within a library system, such as a web-scale discovery index. Of course, discovery layers are not entirely a new concept within libraries—many discovery layer interfaces are currently deployed as end user search interfaces for online public access catalogs. Web-scale discovery services have unique and distinct qualities that make them desirable to libraries: they allow for simple, single searching across the central index; they are responsive and quick; they typically offer a variety of methods for refining search results, such as facets and sorting capabilities; they provide relevance ranked search results; and they allow the user to connect directly with full text whenever appropriate.

In addition to their robust search capabilities that empower the concept of discovery for the end user, web-scale discovery services exhibit an extremely powerful system for indexing an amazing array of content, regardless of where the service is hosted. In addition to integrated library system records, a web-scale discovery service can index a library's digital collections, content from institutional repositories, in addition to other locally created and hosted digital content. Increasingly, this allows libraries the opportunity to uncover for users "hidden" or deep content that may otherwise go undiscovered in general searching. Further, preindexed and remotely hosted content, including e-books, full-text journals, article abstracts, and open access content, the sum of which can seem astronomical to many libraries, can vastly improve resource discovery by the end user in groundbreaking ways. In short, web-scale discovery services provide a unique, seamless, and rapid method for discovery and delivery of relevance ranked content from a vast and rich index.

It should not be surprising that web-scale discovery has evolved (and continues to do so) out of a complex online networked environment where users have become accustomed to a simple, Google-like search functionality that attempts to satisfy search needs in rapid succession across an almost limitless volume of information. Libraries continue to struggle with a perceptibly impossible task—either being a tertiary outpost for users seeking information beyond web-based search engines or simply not being of perceived value to end users in an increasingly connected, online sphere. Where libraries are positioned best to utilize web-scale discovery services,

perhaps, lies in the ability to employ (or even develop) a discovery tool that can significantly increase the use (and discoverability) of the content that has been acquired, licensed, or purchased at great and ongoing costs. Competing interests of end users (e.g., using any information that is most easily discovered (such as through Google) regardless of quality, reliability, and other accepted benchmarks of valuation) and a desire for seamless, rapid search results have necessitated the introduction of such discovery tools across libraries of all types.

CLOSING THE CONTENT GAP

As discovery tools have evolved over the past decade, they have greatly improved access to library collections, connecting users not just with deeper levels of library content, but also incorporating many of the features and tools that users have come to expect with open web-based resources, such as much-improved relevance ranked search results across disparate content formats as well as the look and feel of online portals that are frequently used by the general public. Increasingly, end user expectations of what such discovery tools can actually do is in direct contrast with both capability and compatibility. This, in its primacy, is a discussion of mitigating the content gap already present in all discovery tools. While the fact remains that no discovery service can quite meet the unrealistic expectations of end users (i.e., to discover anything, anytime, instantly), there is much that can be done to increase available content, across all platforms, within discovery services.

There are two significant points that comprise moving positively in this direction. One is, as Marshall Breeding suggests, "a matter of business decisions and strategies," while the other hinges upon libraries' subject expertise and inherent ability to link the user with credible, authoritative resources that exist outside of the immediate purvey of the web-scale discovery index, especially content that is openly accessible on the web but may not be properly indexed within the discovery tool itself (2012, p. 29). Regarding the former, many publishers exert strict control over content, preferring that it be made available through proprietary discovery systems where search result rankings and placement of content within such queries can be manipulated at their own discretion. It would seem obvious, then, that without cooperative efforts amongst all content providers in providing and sharing access to information, discovery tools by and large are

powerless in discovering and accessing a significant volume of information. Further, not all databases and information resources play universally well with discovery tools, requiring a researcher to be aware of resources that exist outside of the discovery tool's capabilities in order to perform a more exhaustive search.

To further compound the matter, open web collections, such as the Victorian Women Writers Project, created by and hosted at Indiana University, might only be discoverable by an end user if the content were either indexed by the discovery tool or the user already possessed a general knowledge of the resource (or, counter to the implicit nature of a discovery system, the user simply employed an open web-based search tool). It is a truism that discovery tools are necessarily limited by their lack of comprehensiveness and that "it's not until discovery services truly provide access to a comprehensive representation of all the library's collection components that they can achieve their true potential" (Breeding, 2012, p. 29). Of course, this is a rather simple observation to make. Common sense would dictate that comprehensiveness within a discovery system is an enormous undertaking. The Open Discovery Initiative, a working group of the National Information Standards Organization, has been established to define best practices and standards for index-based library discovery services and to provide transparency in the ways that content is represented. Such standards and practices would prove essential to libraries in evaluating discovery systems that are best suited given the library's unique collections.

ADDING VALUE TO DISCOVERY

Significant progress has been made in recent years as more libraries have adopted discovery systems to their cache of tools aimed at increasing discoverability of resources for the end user. However, there are many improvements to be made (and looked forward to in future) that will serve to add value to the discovery process. To ensure access and discoverability of resources within the library's collections, inside its discovery system index, and across the open web, libraries must provide adequate content description through descriptive and subject cataloging, consider the organization of resources within the greater collection (via classification schema and indexing services), and make general provisions for access over time (preservation, archiving, etc.). Tangentially, the presentation of search results

within a discovery system directly influences the system's overall effectiveness. An end user will naturally want the best (and most relevant) results to appear first, and it is a logical extension of thought that search relevancy methodology will need continuous improvement across all discovery systems to meet user needs (and satisfy library expectations).

As discovery systems work on such improvements, considerations for relevancy of web-based searching will likely serve to facilitate this evolution. Marshall Breeding suggests that "library search will also benefit from a more personalized approach. Information such as the user's academic department and previous search history can be great clues regarding the kinds of materials that would be considered most relevant in the user's search results" (2012, p. 30). Additional concerns of a discovery system's ability to allow for proper browsing of the library's collection, as well as e-book discoverability through such tools, remain at the forefront of continued improvements to discovery services.

INTEGRATING TRADITIONAL TOOLS WITH DISCOVERY SYSTEMS

Traditionally, libraries facilitated the process of discovery for their patrons through a variety of means previously discussed in this chapter—the catalog, personalized reference service, bibliographies, and research guides. The need and desire for such tools has not significantly waned in the age of discovery; however, the inherent value of the traditional pathfinder, even when translated for web-based environments, is not always fully realized. Discovery system vendors and developers have recognized the need to provide access, either directly or indirectly (via linking services), to such tools and resources that are not inherently indexed within the service itself. Subject-based research guides (or pathfinders), FAQs or knowledge base articles, general help pages, librarian profiles, and other information about library outreach and programmatic events would all be potentially useful content if woven into the discovery system. Adding localized content to a web-scale discovery system is, theoretically, an easier task to accomplish than adding content from competing publishers and indexes. One such endeavor in this area is the integration of Springshare's LibGuides into Serials Solutions' Summon service. Summon's application programming interface (API) allows for more thorough integration and interplay between the two systems, giving end users more options for how they discover information by

and about the library. Similarly, the EBSCO Discovery Service allows for the provision of links to frequently used content sources—such as LibGuides or other relevant library information, such as operating hours or general announcements—via specialized widgets that can be added at various levels within the system. Increasingly, libraries will see the need to add this additional content to discovery services to improve discoverability by end users.

DISCOVERY TOOL USE

To be certain, dramatic shifts in technology in the digital age, as well as responsive and meaningful approaches to such changes, have radically altered how users interact with information. Search engines such as Google, social media networks like Twitter and Facebook, and an overall cultural change have directed libraries to employ tools that work in a very similar fashion to web-based search engines due to expediency, convenience, and, perhaps most of all, ease of use. Many studies have reached the conclusion that in order for libraries to exact change and remain relevant in the 21st century, users must engage with the library quite similarly to how they engage with the online world. Such familiarity appears important to users, who through rote practice and the ubiquity of online networks in their daily lives have become not only confident in their daily practice, but also wholly comfortable with how information is presented to them within a web-based environment outside of the library. The Google search interface, a simple, single box, has become the de facto standard for information gathering amongst most undergraduate college students, for example, and has created expectations, unrealistic or not, that engaging with library tools and resources should replicate that familiarity and comfort.

The reality that libraries are faced with, though, suggests the traditional library arrangement of resources, especially via the library website, as well as the inherent complexity of library catalogs, databases, and indexes, create undue anxiety for users and serve to counteract any implied value that such tools might provide in the way of help. Counter to the argument by some librarians that discovery tools serve to "dumb down students' information search skills," it has been found that a "simpler and more direct way of information retrieval would actually free up time for instruction librarians to teach about information itself and how to engage with it in a useful way, rather than teaching the 'click here, click there' procedural steps and

Boolean search strategies which students are unlikely to use again in the future" (Cmor & Li, 2012, p. 1). Considering user search behavior, then, it would be reasonable to suggest discovery tool use is typically governed by a variety of trends already inherent (and ingrained) in user habits: users typically engage with information through resources outside of the library (e.g., the open web); users have the expectation that resource discovery and immediate availability (delivery) exist in tandem with each other; and with increasing frequency, users engage with nontraditional information sources.

It is true that users do not typically view the library as a starting point for information gathering. Most end user surveys have found that an overwhelming majority of users begin a search for information using an open web search engine such as Google. However, these impediments should not discourage libraries wishing to adopt discovery systems. Rather, they should serve as reminders that a user-centered approach that is considerate of the needs of its community of users, while remaining customized to its unique collections and strengths, will serve to facilitate resource discovery with increased satisfaction and confidence in much the same way that has become customary outside of the library environment. A recognition by libraries that Google has significantly altered user search expectations (and habits) has resulted in the increased use of resource discovery tools, and as the tools themselves become more sophisticated (improved relevancy ranking, etc.), it is becoming evident that they may well provide the search solution that is the best compromise between user expectation and realistic delivery of information. As discovery systems evolve and improve, libraries must continually reassess how their catalogs and information systems keep pace with user expectation and understanding. Some catalogs and databases perform certain search and retrieval tasks implicit in the notion of discovery far better than current discovery systems. Libraries will need to ascertain how best to leverage a discovery tool within its existing structure to best enable the end user's ability to uncover and access the most appropriate resources for his or her needs.

FUTURE DIRECTIONS: OPEN-SOURCE AND COMMERCIAL AVENUES FOR DISCOVERY

A survey of current literature will reward the searcher with countless articles about choosing the right discovery tool by using a whole host of criteria,

including customizability, interoperability, and price. There are seemingly numerous options for libraries of all sizes and operating budgets, with an equal weight of professional commentary both for and against almost any solution. The main players in the commercial market for discovery systems were previously discussed, and it is worth mentioning that ongoing development in this arena will produce new contenders in the market, some of which stem from the same vendors mentioned. Ex Libris's Alma, billed as a next-generation library services framework, aims to offer a "suite" of operational functions that address selection, acquisition, digitization, and management, among other things, which can be integrated with external systems. Innovative Interface's Sierra Services Platform promises a similar set of features with open development and customization. OCLC's WorldShare Management Services is self-described as the first cooperative management service for libraries. It is a cloud-based system with enhanced end user discovery with "Google-like" searching. It is likely that many libraries may transition to these types of platforms in future, moving away from legacy integrated library systems. Alternatively, Kuali OLE (Open Library Environment), which identifies itself as the "first system designed by and for academic and research libraries for managing and delivering intellectual information," is an enterprise-ready, community-source software package aimed at managing and providing access to collections and licensed and local digital content.

Such library services platforms, incorporating discovery tools with management and delivery systems, may be seen as the next evolution of the library catalog. While next-generation catalogs do not typically fit within the outline we have formed to define resource discovery tools, they certainly present the next trend in a mix-and-match approach to library automation. Importantly, there are many open-source discovery interfaces that are being deployed successfully across library systems. The Indiana University Bloomington Libraries, for example, have combined the Blacklight catalog (an open-source Ruby on Rails engine that provides a discovery interface for Apache Solr) with the EBSCO Discovery Service. Similarly, the University of Virginia has combined its own Blacklight catalog with Ex Libris's Primo Central Index. Villanova University's VuFind catalog (an open-source library resource portal developed for and by libraries) has Serials Solutions' Summon service as its incorporated discovery

layer. Other examples of open-source discovery systems include: OpenBib, a search portal currently under development that is customizable and extendable; eXtensible Catalog, which comprises four software components to provide end user discovery that works well with Drupal (an open-source content management platform); SOPAC (Social Online Public Access Catalog), a module that integrates the library catalog system with the Drupal content management system while allowing users to tag, rate, and review collections holdings, which are then incorporated into the discovery index, effectively creating a "community-driven catalog system"; and, Xerxes, a mobile-ready library portal that is customizable and provides citation management and integration features.

WHITHER LIBRARIES IN THE AGE OF DISCOVERY?

Information scientist and theorist Frederick Wilfrid Lancaster, whose own work typically centered on online retrieval, envisioned, in 1978, the future as "one of a society whose formal communication will be paperless" and "as a consequence, library problems in the long term do not relate to inadequate space or even to inadequate financial resources. They all come down to one problem only: justification for existence, simple survival" (Thompson, 1982, p. 109). Lancaster asked broadly, "can libraries . . . survive in a largely electronic world?" (p. 109). For his own part, he did try to answer this question throughout his career, but he suggested (one year later in 1979) that libraries in the year 2000 "will with only very few exceptions offer 'multisource' catalogues" that include "not only entries for all the materials held by the network or networks to which a particular library belongs, but also entries for all externally accessible databases, primary and secondary, which any member library chooses to include" (p. 110). Certainly, Lancaster may have been a bit quick to presume the state of libraries in the year 2000; however, the current development and evolution of discovery tools available to libraries may very well serve to facilitate Lancaster's wishful prognostication in the not-too-distant (and collaborative) future.

REFERENCES

Breeding, M. (2014). Web-scale discovery services. *American Libraries*, *45*(1/2), 25.

Breeding, M. (2012). Looking forward to the next generation of discovery services. *Computers in Libraries*, *32*(2), 28–31.

Cmor, D., & Li, X. (2012). Beyond Boolean, towards thinking: Discovery systems and information literacy. *Proceedings of the IATUL Conferences*. Retrieved from http://docs.lib.purdue.edu/iatul/2012/papers/7

Dempsey, L. (2012). Thirteen ways of looking at libraries, discovery, and the catalog: Scale, workflow, attention. *Educause Review Online*. Retrieved from http://www.educause.edu/ero/article/thirteen-ways-looking-libraries-discovery-and-catalog-scale-workflow-attention

Lynch, C. A. (1995). Networked information resource discovery: An overview of current issues. *IEEE Journal on Selected Areas in Communications, 13*(8), 1505–1522. http://dx.doi.org/10.1109/49.464719

Ranganathan, S. R. (1931). *The five laws of library science*. Madras: The Madras Library Association.

Thompson, J. (1982). *The end of libraries*. London: Clive Bingley Limited.

Vaughn, J. (2011). Web scale discovery: What and why. *Library Technology Reports, 47*(1), 5–11.